GEODEZJA I KARTOGRAFIA
GEODESY AND CARTOGRAPHY
Vol. 58, No 1, 2009, pp. 3-21

M_{split} estimation. Part I: Theoretical foundation

Zbigniew Wiśniewski

Institute of Geodesy, University of Warmia and Mazury
1 Oczapowskiego St., 10-957 Olsztyn, Poland
e-mail: zbyszekw@uwm.edu.pl

Received: 27 November 2008/Accepted: 27 March 2009

Abstract: The present paper consists of two parts. The first part presents theoretical foundations of M_{split} estimation with reference to the previous author's paper (Wiśniewski, 2009). This time, some probabilistic assumptions are described in detail. A new quantity called f-information is also introduced to formulate the split potential in more general way. The main aim of this part of the paper is to generalize the target function of M_{split} estimation that is the basis for a new formulation of the optimization problem. Such problem itself as well as its solution are presented in this part of the paper.

The second part of the paper presents some special case of M_{split} estimation called squared M_{split} estimation (also with reference to the mentioned above paper of the author). That part presents a new solution and development in the theory of this version of M_{split} estimation and some numerical examples that show properties of the method and its application scope.

Keywords: Geodetic adjustment, M-estimation, split potencial of an observation set

1. Introduction

The maximum likelihood method (ML-estimation) is one of the important estimation methods and it is also applied in geodetic computations (e.g. Grodecki, 1999; Koch, 1986; Wiśniewski, 1987; Yu, 1996). Let $v_i = y_i - \theta$ be a functional model of geodetic observations y_i, $i = 1, \ldots, n$, where θ is an unknown parameter and v_i is a standardized measurement error (without loosing generality and to simplify notation the variable v_i is assumed to be standardized also in the next parts of the paper). A quantity $\hat{\theta}$ is the ML-estimate of a parameter θ of a random variable $\mathbf{y} = [y_1, \ldots, y_n]^T$ with the density function $f(\mathbf{y};\theta)$ when

$$\min_{\theta} \varphi(\mathbf{y};\theta) = \varphi(\mathbf{y};\hat{\theta}) \tag{1}$$

where

$$\varphi(\mathbf{y};\theta) = -\sum_{i=1}^{n} \ln f(y_i;\theta) \tag{2}$$

(the random variables are assumed to be mutually independent).

However, it is M-estimation that has played a major role in geodetic elaborations for some years. Foundations of this estimation method were formulated and presented by Huber (1964, 1981) under the assumption that the term $-\ln f(y;\theta)$ inside the target function Eq. (2) can be replaced with another arbitrary function $\rho(y;\theta)$. Thus, $\hat{\theta}$ is an M-estimate of the parameter when

$$\min_{\theta}[\varphi(\mathbf{y};\theta) = \sum_{i=1}^{n}\rho(y_i;\theta)] = \sum_{i=1}^{n}\rho(y_i;\hat{\theta}) \tag{3}$$

For that reason, M-estimation class contains for example the least squares method (LS-method) with the function $\rho(y;\theta) = v^2$ and the principle of the alternative choice (Kadaj, 1984) with the function $\rho(y;\theta) \propto f(y;\theta)$ ($\propto$ is a proportionality sign). The last of these methods is robust against gross errors. The estimates of the parameter θ obtained by applying this method or other robust estimates (in a sense of this paper interest) will be denoted as $\hat{\theta}_R$. Such option in choosing functions $\rho(y;\theta)$ results in a broad class of robust M-estimates. Usually, such estimates are based on the following probabilistic model of a gross error (ε-contaminated)

$$P_\theta = (1-\varepsilon)P_{\theta_\alpha} + \varepsilon P_{\theta_\beta} \qquad (0 \le \varepsilon \le 1) \tag{4}$$

where P_{θ_α} is a probability distribution that belongs to the family $\mathcal{P}_\alpha = \{P_{\theta_\alpha} : \theta_\alpha \in \Theta_\alpha\}$ of all acceptable probability distributions (for random variables y) and P_β is a "strange" distribution that belongs to another family $\mathcal{P}_\beta = \{P_{\theta_\alpha} : \theta_\beta \in \Theta_\beta\}$ (e.g. Serfling, 1980). The $\mathcal{P}_\alpha$ and $\mathcal{P}_\beta$ families are indexed with parameters $\theta_\alpha \in \Theta_\alpha$ and $\theta_\beta \in \Theta_\beta$, respectively. Huber's function is a good example of $\rho(y;\theta)$ function that refers to the model Eq. (4) (Huber, 1981; Hampel et al., 1986)

$$\rho(y;\theta) = \begin{cases} \rho_\alpha(y;\theta) = v^2 & for\,|v| \le t \Leftrightarrow y \sim P_{\theta_\alpha} \\ \rho_\beta(y;\theta) = t\,|v| - \dfrac{1}{2}t^2 & for\,|v| > t \Leftrightarrow y \sim P_{\theta_\beta} \end{cases} \tag{5}$$

where $t > 0$ (e.g. $t = 2.0, 2.5, 3.0$). There are also other methods where $\rho_\beta(y;\theta)$ function is two-segmented (e.g. Hampel et al., 1986; Yang, 1991, 1994, 1999).

Geodetic applications of robust M-estimation are presented in many papers for example Krarup and Kubik (1983), Huang and Mertikas (1995), Koch (1996), Gui and Zhang (1998), Yang (1999), Götzelmann et al. (2006). Generalization of M-estimation to the case of dependent observations can be found in Xu (1989), Yang (1994) and Yang et al. (2002). The principles of M-estimation were also applied to develop robust Kalman filter (e.g. Koch and Yang, 1998; Yang et al., 2001), robust collocation (Yang, 1992), robust estimation of variance coefficient (Wiśniewski, 1999) or robust estimation of variance-covariance matrices (Yang, 1997). Theoretical properties of M-estimation can be analyzed by using the influence function (e.g. Hampel, 1974; Serfling, 1980; Hampel et al., 1986) or the breakdown point (e.g. Rousseeuw, 1984; Hampel et al., 1986; Xu, 2005). It is well known that robust M-estimation results in decreasing of influence of outliers on the final estimate $\hat{\theta}_R$ (except for some critical case presented in (Xu, 2005) or, from the other point of view, in (Prószyński, 1997, 2000)). Let $\hat{\theta}_R$ be the most robust estimate, i.e. the estimate that is free of the influence of all outliers of unacceptable distribution P_{θ_β} (outliers are rejected during the estimation process). In such theoretical and idealistic case, the estimate $\hat{\theta}_R$ of the parameter θ is the estimate of the parameter θ_α at the same time. It is obvious that observations rejected during the estimation process are not interesting from the M-estimation point of view. Furthermore, the parameter θ_β of unacceptable distribution P_{θ_β} is not estimated at all. It is reasonable if outliers results from gross errors. However, even then it would be interesting to compute the estimate $\hat{\theta}_\beta$ to help understand or analyze properties or sources of gross errors. On the other hand, outliers are not always effects of faults, slips etc. Observation sets can sometimes contain some observations with deterministic errors or systematic ones that are hard to identify. Also, if observation sets that were measured at two different epochs are elaborated together, some observations for example concerning some displaced points can be regarded as outliers. In such cases or other similar ones (like a disturbed laser scanning) an estimation of the parameter θ_β, not only θ_α, seems reasonable. Such estimation process can be carried out by applying M_{split} estimation.

Theoretical foundation of the method proposed in (Wiśniewski, 2009) and called M_{split} estimation is a general assumption that each observation described by the functional model $v = y - \theta$ can have either of two distributions $P_\alpha \in P_{\theta_\alpha}$ or $P_\beta \in P_{\theta_\beta}$. Consequently, the model $v = y - \theta$ is naturally split into two competitive ones $v_a = y - \theta_\alpha$ and $v_\beta = y - \theta_\beta$. In the cited paper, it is assumed that every observation y has got some potential describing the chance to identify the observation with either of two possible distributions P_{θ_α} or P_{θ_β}. Such potential, called the elementary split potential, was referred to the probability $p_\alpha\ (y;\ \theta_\alpha) \Leftrightarrow y \sim P_{\theta_\alpha}$ and information $I_\beta(y;\ \theta_\beta) = -\ln\ p_\beta(y;\ \theta_\beta)$ (e.g. Jones and Jones, 2000), that every observation y can provide under competitive assumption that $y \sim P_{\theta_\beta}$. Generally, one can say that, according to the presented theory, quantities $\hat{\theta}_\alpha$ and $\hat{\theta}_\beta$ are M_{split} estimates of competitive parameters θ_α and θ_β if only they maximize the split potential.

The next section of the paper presents assumptions and further theoretical development of M_{split} estimation in detail. Theory is referred to the functional model of geodetic observations. The general target function of M_{split} estimation as well as conditions for respective optimization problem solution are also presented.

2. Fundamental assumptions

2.1. Idea of M_{split} estimation

Let the following

$$\{1.1, 1.3, 1.4, 1.5, 1.7, 3.4, 3.5, 3.6\}$$

be a set that contains realizations of a random variable Y. Such realizations regarded as observations can be referred to the following functional model

$$v_i = y_i - \theta, \quad i = 1, ..., 8$$

One can estimate the parameter θ by applying the LS-method and obtain $\hat{\theta}_{LS} = 2.19$. However, the set elaborated contains some outliers. Thus an ε – contaminated distribution $P_\theta = (1 - \varepsilon)P_{\theta_\alpha} + \varepsilon P_{\theta_\beta}$ can be assumed as a probabilistic model of the variable Y. Taking such assumption the parameter θ can now be estimated by using a method that belongs to the robust M-estimation class. Thus if the weight function defined as follows (e.g. Kadaj 1988; Yang, 1991; Huang and Mertikas, 1995; Gui and Zhang, 1998)

$$w(v) = \frac{\partial\rho(y;\theta)}{\partial(v^2)} \tag{6}$$

and the Danish method (e.g. Krarup and Kubik, 1983; Zhong, 1997), where

$$w(v) = \begin{cases} 1 & for\,|v| \leq t \\ \exp\{-l(v-t)^2\} & for\,|v| > t \end{cases} \tag{7}$$

is applied, then one can compute the Danish M-estimator as $\hat{\theta}_R = 1.49$ (under assumptions $l = 0.5$, $t = 0$). Very similar results can be obtained by using other robust methods or the Danish method but under different assumptions concerning the control parameters l and t. Regardless to the method choice, the influence of the outliers 3.4, 3.5, 3.6 on the robust estimate of the parameter θ is decreased or even eliminated. This common property results from the assumption that the parameter θ_α should lie near to the densest observation concentration. Thus, in robust M-estimation, $\rho(y;\, \theta)$ functions are designed in such way to fulfill this postulate. For example, Wiśniewski (1993) proposed that $\rho(y;\, \theta) = -\ln f^{RP}(y;\, \theta)$ where $f^{RP}(y;\, \theta)$ is the density function of an asymmetric distribution P_θ that belongs to the family of the Pearson distributions (e.g. Elderton, 1953). For the method called the principle of the alternative choice, Kadaj (1984) assumed that $\rho(y;\, \theta) \propto -f^{ND}(y;\, \theta)$ where $f^{ND}(y;\, \theta)$ is the density function of a distribution P_θ that belongs to the family of normal distributions. In this two cases, the weight function Eq. (6) is a continuous and concave function with the maximum at the point $v = 0$ (for example for the second presented method $w(v) \propto \exp(-v^2)$, e.g. Kadaj, 1984). Thus, such weight functions assign the biggest values of equivalent weights to observations, which values, are close to the parameter θ_α. For some methods that belong to the M-estimation class and that are based on the model Eq. (4) such requested properties of the weight function Eq. (6) are obtained by combining the function $\rho_\alpha(y;\, \theta) = v^2$ with $\rho_\beta(y;\, \theta)$ where:

$$w_{\alpha}(y;\theta) = \partial\rho_{\alpha}(y;\theta)/\partial(v^2) = 1$$
$$w_{\beta}(y;\theta) = \partial\rho_{\beta}(y;\theta)/\partial(v^2), \quad \forall \left|v_j\right| > |v_i| : w_{\beta}(y_j;\theta) < w_{\beta}(y_i;\theta)$$

(see Eq. (5), Eq. (7) and, e.g. Hampel et al., 1986). In all these cases, the main task of the weight function $w(v)$ (or its part $w_{\beta}(v)$) is to assign small values of equivalent weights to the observations that are supposed to be distributed according to the "strange" distribution P_{θ_β}.

Let us consider the observation set presented above one more time. One can suppose that since $\hat{\theta}_R = 1.49$ then the influence of observations 3.4, 3.5, 3.6 on this value was rather small actually. Let us pay attention to the fact that these observations are concentrated around another parameter, for example θ_β. As it was assumed earlier the observation set contains realizations of two different random variables Y_α and Y_β with different parameters θ_α and θ_β. However, we were only interested in the parameter θ_α estimated by $\hat{\theta}_R$. Such approach to estimation of parameters seems doubtful if we know that an observation set contains realizations of two random variables. Since two parameters θ_α and θ_β can be assigned to the observation set it is natural to estimate them both. In some cases knowledge of the estimate $\hat{\theta}_\beta$ value can be insignificant or inutile but it does not mean that estimation of both parameters together does not influence the value of the parameter $\hat{\theta}_\alpha$. One should expect that during such integrated process of estimation, estimates $\hat{\theta}_\alpha$ and $\hat{\theta}_\beta$ will "attract" observations that suit them best, respectively. Thus the estimate $\hat{\theta}_\alpha$ can be a function of realizations of the variable Y_α only (that is not always possible for robust estimates $\hat{\theta}_\alpha$). In other practical cases, that were presented in the section 1, both values $\hat{\theta}_\alpha$ and $\hat{\theta}_\beta$ seem interesting. The graphical interpretation of the present observation set as well as the values of the estimates $\hat{\theta}_{LS}$ and $\hat{\theta}_R$ is presented in Figure 1a. The same figure presents also the desired places for the estimates $\hat{\theta}_\alpha$ and $\hat{\theta}_\beta$. The similar interpretation but for the observation set with only one outlier is presented in Figure 1b.

Traditional approach to calculation of the estimates $\hat{\theta}_\alpha$ and $\hat{\theta}_\beta$ assumes some arbitrary principal of assigning realizations of the random variable Y to either of these two estimators. M_{split} estimation makes two competitive, objective assumptions (the observation set is not divided into parts):

- observations are realizations of the random variable $Y_\alpha \sim P_{\theta_\alpha}$
- observations are realizations of the random variable $Y_\beta \sim P_{\theta_\beta}$

Such assumptions result in split of the probabilistic model of observations (every observation can be described by two competitive distributions) but also lead to two competitive functional models (this fact is discussed in the next parts of the paper). During the estimation process either of models becomes more realistic, probable for each observation, i.e. it suits respective observation better (one can say it "wins" the competition). One can suppose that a single observation can provide some potential that can give opportunity to identify it with either of two competitive models (identifying on the whole or only in part, i.e. an observation can be identified with the both models).

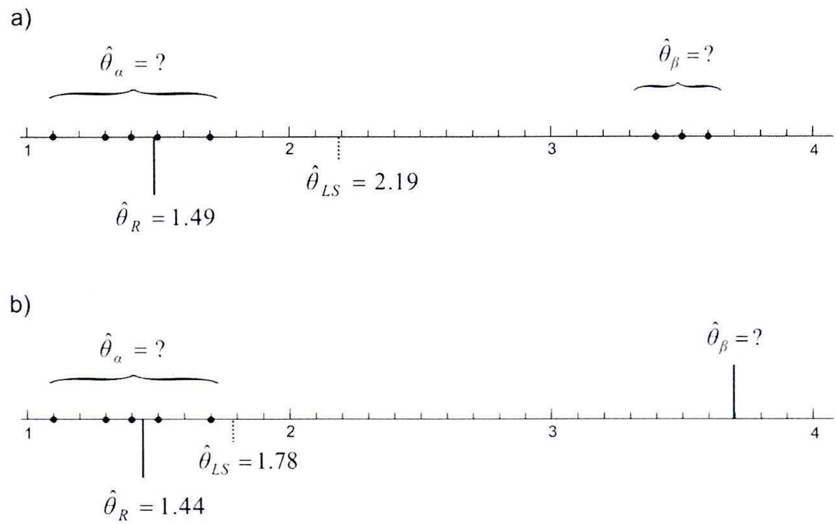

Fig. 1. Graphical interpretation of M_{split} estimation idea

2.2. Elementary split potential

Let us consider a single observation y that is a realization of a random variable Y with the following ε-contaminated distribution

$$P_\theta = (1-\varepsilon)P_{\theta_\alpha} + \varepsilon P_{\theta_\beta}$$

Let there be a number that can describe the chance to identify the observation with either of two distributions P_{θ_α} or P_{θ_β}. Let these distributions be discrete ones with the probability functions $p_\alpha(y;\ \theta_\alpha)$ and $p_\beta(y;\ \theta_\beta)$. It means that for each observation there are two probabilities $p_\alpha(y;\ \theta_\alpha) = p_\alpha$ and $p_\beta(y;\ \theta_\beta) = p_\beta$. Then the number mentioned above (and obtaining values from the interval $< 0, 1 >$) can be written as the following quantity (Wiśniewski, 2009)

$$K_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) = p_\alpha(y;\theta_\alpha)\uparrow I_\beta(y;\theta_\beta) = p_\beta(y;\theta_\beta)\uparrow I_\alpha(y;\theta_\alpha) \tag{8}$$

where $a \uparrow b = a^b$. The term $I(y;\ \theta) = -\ln\ p(y;\ \theta)$ is an amount of information that every single observation can provide if the probability of the observation appearance is $p(y;\ \theta) = p$ (e.g. Jones and Jones, 2000). The quantity $K_{\alpha,\beta}(y;\ \theta_\alpha, \theta_\beta)$ is called the elementary split potential (Wiśniewski, 2009). Let $p_\beta = p_\beta(y;\ \theta_\beta) = 1$, then the amount of information that the observation can provide is equal to zero

$$I_\beta(y;\theta_\beta) = -\ln p_\beta(y;\theta_\beta) = 0$$

Because $p_\beta = 1$ thus such observation was expected, exactly. Its appearance did not provide any new information. In such case, regardless of the probability value

$p_\alpha = p_\alpha(y, \theta_\alpha)$, the split potential obtains its biggest value $K_{\alpha,\beta}(y;\ \theta_\alpha,\ \theta_\beta) = 1$ (it results from the first part of Eq. (8)). The split potential value equals to 1 also when $p_\alpha(y;\ \theta_\alpha)$, regardless of the value of $p_\beta(y;\ \theta_\beta)$. The split potential approaches zero when $p_\alpha(y;\ \theta_\alpha) < 1$ and $p_\beta(y;\ \theta_\beta) \rightarrow 0$ at the same time. Of course, one can consider the second part of Eq. (8) and discuss the situation equivalently on the base of the information $I_\alpha(y;\ \theta_\alpha)$. The split potential proposed in the mentioned paper for discrete distributions has the following general properties

$$
\begin{aligned}
&K_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) = 1 \quad when: \\
&\qquad I_\beta(y;\theta_\beta) = 0 \Leftrightarrow p_\beta(y;\theta_\beta) = 1 \ \ (regardless\ \ of\ \ p_\alpha(y;\theta_\alpha)) \\
&\qquad \text{or} \\
&\qquad I_\alpha(y;\theta_\alpha) = 0 \Leftrightarrow p_\alpha(y;\theta_\alpha) = 1 \ \ (regardless\ \ of\ \ p_\beta(y;\theta_\beta))
\end{aligned}
$$

and

$$
\begin{aligned}
&K_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) \rightarrow 0 \quad when: \\
&\qquad p_\beta(y;\theta_\beta) \rightarrow 0 \Leftrightarrow I_\beta(y;\theta_\beta) \rightarrow \infty \quad and \quad p_\alpha(y;\theta_\alpha) < 1 \\
&\qquad \text{or} \\
&\qquad p_\alpha(y;\theta_\alpha) \rightarrow 0 \Leftrightarrow I_\alpha(y;\theta_\alpha) \rightarrow \infty \quad and \quad p_\beta(y;\theta_\beta) < 1
\end{aligned}
$$

The elementary split potential idea can be generalized to the case of random variable with continuous distributions P_0 that belong to the family $\mathcal{P} = \{P_0 > f(y;\ \theta) : \theta \in \Theta\}$ where $f(y;\ \theta)$ is a density function (the sign $>$ means that this function is assigned to the distribution P_θ). For such purpose, the f-information defined as follows

$$I^f(y;\theta) \propto -\ln cf(y;\theta), \quad c > 0 \tag{9}$$

is now introduced. For example, if y is normally distributed and its distribution belongs to the normal distribution family $\mathcal{P} = \{N[\theta, \sigma] : \theta \in \theta\}$ then

$$I^f(y;\theta) = (y-\theta)^2 = v^2 \quad for \ \ c = 2\sqrt{\pi} \tag{10}$$

Let us pay attention to the fact that the amount of the f-information provided by the observation $Y = y$ grows larger with the increase of the absolute value of the random error $v = y - \theta$ (Fig. 2a). The maximum probability of appearance is assigned to errors close to zero. Thus if θ_α is a parameter of the variable Y distribution and if an observation y with a big and less probable error appears then the information that y can provide is rather big (Fig. 2a). Such information can even point at the fact that the parameter θ_α is not a parameter of the y distribution and should be replaced with for example θ_β (it is very important fact in $M_{\rm split}$ estimation theory). On the other hand, the amount of f-information about the parameter θ_β can be close or even equal to zero (Fig. 2b).

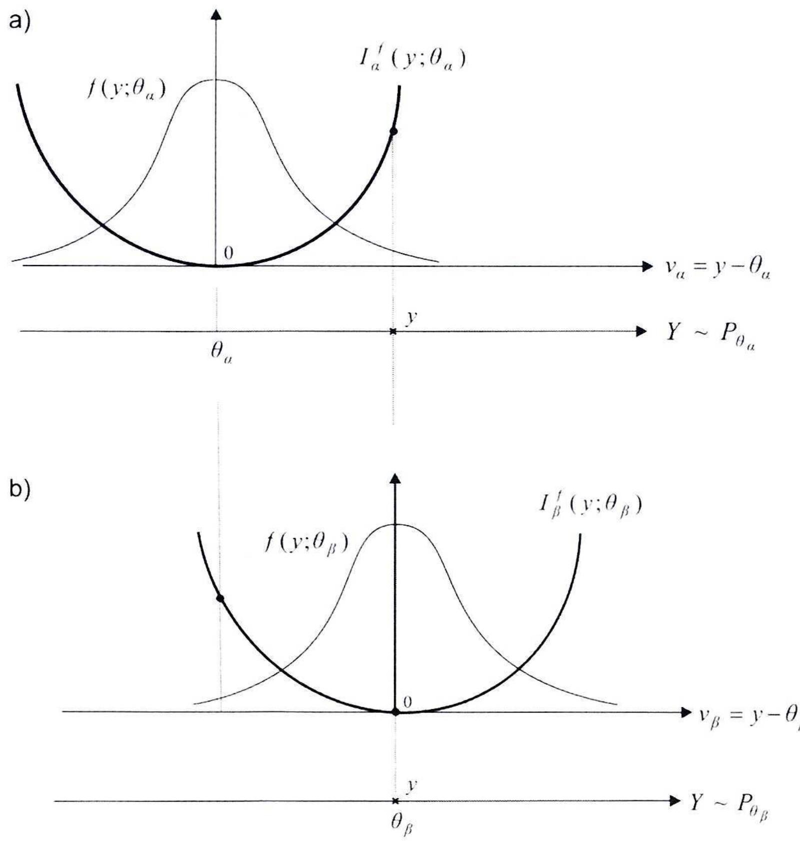

Fig. 2. Density function and f-information

The elementary split potential Eq. (8) can be rewritten in a new following form by using the density function and f-information

$$K_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) = c_1 f_\alpha(y;\theta_\alpha) \uparrow I_\beta^f(y;\theta_\beta) = c_2 f_\beta(y;\theta_\beta) \uparrow I_\alpha^f(y;\theta_\alpha) \tag{11}$$

Taking $I^f(y;\,\theta) = -\ln cf(y;\,\theta)$, the elementary split potential for the observation y can be written as follows (for $c_1, c_2 > 0$)

$$\begin{aligned} K_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) &= c_1 f_\alpha(y;\theta_\alpha) \uparrow [-\ln c_2 f_\beta(y;\theta_\beta)] = \\ &= c_2 f_\beta(y;\theta_\beta) \uparrow [-\ln c_1 f_\alpha(y;\theta_\alpha)] \end{aligned} \tag{12}$$

The potential Eq. (12) is based on the assumption that the observation y can be a realization of either of two competitive random variables $Y_\alpha \sim P_{\theta_\alpha}$ or $Y_\beta \sim P_{\theta_\beta}$ where P_{θ_α} and P_{θ_β} are probability distributions that belong to the following families, respectively:

$$\mathcal{P}_\alpha = \{P_{\theta_\alpha} > f_\alpha(y;\theta_\alpha) : \theta_\alpha \in \Theta_\alpha\},\ \ \mathcal{P}_\beta = \{P_{\theta_\beta} > f_\beta(y;\theta_\beta) : \theta_\beta \in \Theta_\beta\}$$

The natural logarithm of the split potential plays a very special role in M_{split} estimation. Thus taking

$$k_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) = \ln K_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) \tag{13}$$

one can write

$$\begin{aligned} k_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) = \ln K_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) = & \\ &= \ln\{c_1 f_\alpha(y;\theta_\alpha) \uparrow [-\ln c_2 f_\beta(y;\theta_\beta)]\} = \\ &= \ln\{c_2 f_\beta(y;\theta_\beta) \uparrow [-\ln c_1 f_\alpha(y;\theta_\alpha)]\} = \\ &= -\ln c_1 f_\alpha(y;\theta_\alpha) \ln c_2 f_\beta(y;\theta_\beta) \end{aligned} \tag{14}$$

Since

$$\begin{aligned} -\ln c_1 f_\alpha(y;\theta_\alpha) &= I^f_\alpha(y;\theta_\alpha) \\ -\ln c_2 f_\beta(y;\theta_\beta) &= I^f_\beta(y;\theta_\beta) \end{aligned}$$

then

$$k_{\alpha,\beta}(y;\theta_\alpha,\theta_\beta) = -I^f_\alpha(y;\theta_\alpha) I^f_\beta(y;\theta_\beta) \tag{15}$$

2.3. Split potential of observation vector (global split potential)

Let $\mathbf{Y}_\alpha = [Y_{1\alpha}, Y_{2\alpha}, \cdots, Y_{n\alpha}]^{\mathrm{T}}$ and $\mathbf{Y}_\beta = [Y_{1\beta}, Y_{2\beta}, \cdots, Y_{n\beta}]^{\mathrm{T}}$ be random vectors and let $P_{\boldsymbol{\theta}_\alpha}$ and $P_{\boldsymbol{\theta}_\beta}$ be their distributions, respectively. The distributions belong to the following respective families:

$$\mathcal{P}_\alpha = \{P_{\boldsymbol{\theta}_\alpha} \succ f_\alpha(\mathbf{y};\boldsymbol{\theta}_\alpha) : \boldsymbol{\theta}_\alpha \in \boldsymbol{\Theta}_\alpha \quad \mathcal{P}_\beta = \{P_{\boldsymbol{\theta}_\beta} \succ f_\beta(\mathbf{y};\boldsymbol{\theta}_\beta) : \boldsymbol{\theta}_\beta \in \boldsymbol{\Theta}_\beta\} \tag{16}$$

($\boldsymbol{\theta} \in \mathfrak{R}^n$ – parameter vector). Additionally, let random variables $Y_{i\alpha}$, $Y_{i\beta}$, $i = 1, ..., n$, be mutually independent. Then

$$f_\alpha(\mathbf{y};\boldsymbol{\theta}_\alpha) = \prod_{i=1}^{n} f_\alpha(y_i;\theta_{i\alpha}) \tag{17}$$

$$f_\beta(\mathbf{y};\boldsymbol{\theta}_\beta) = \prod_{i=1}^{n} f_\beta(y_i;\theta_{i\beta}) \tag{18}$$

Let $\mathbf{y} = [y_1, y_2, ..., y_n]^{\mathrm{T}}$ be a vector of observations and let it be a realization of either of two competitive random variables $\mathbf{Y}_\alpha \sim P_{\boldsymbol{\theta}_\alpha}$ and $\mathbf{Y}_\alpha \sim P_{\boldsymbol{\theta}_\alpha}$. According to the earlier assumptions, every observation y_i has its own assigned elementary split potential $K_{\alpha,\beta}\,(y_i;\theta_\alpha,\theta_\beta)$ so a proper split potential should also be assigned to the whole vector $\mathbf{y}$. In the paper (Wiśniewski, 2009), such potential was called as the global split potential and defined as follows

$$K_{\alpha,\beta}(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)=\prod_{i=1}^{n}K_{\alpha,\beta}(y_i;\theta_{i\alpha},\theta_{i\beta}) \tag{19}$$

Considering Eqs. (11) and (12) and also Eqs. (16)-(19), the potential $K_{\alpha,\beta}(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)$ can be written in the form

$$\begin{aligned}K_{\alpha,\beta}(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)&=\prod_{i=1}^{n}\{c_1f_\alpha(y_i;\theta_{i\alpha})\uparrow I_\beta^f(y_i;\theta_{i\beta})\}=\\&=\prod_{i=1}^{n}\{c_2f_\beta(y_i;\theta_{i\beta})\uparrow I_\alpha^f(y_i;\theta_{i\alpha})\}\end{aligned} \tag{20}$$

or

$$\begin{aligned}K_{\alpha,\beta}(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)&=\prod_{i=1}^{n}\{c_1f_\alpha(y_i;\theta_{i\alpha})\uparrow[-\ln c_2f_\beta(y_i;\theta_{i\beta})]\}=\\&=\prod_{i=1}^{n}\{c_2f_\beta(y_i;\theta_{i\beta})\uparrow[-\ln c_1f_\alpha(y_i;\theta_{i\alpha})]\}\end{aligned} \tag{21}$$

The logarithmic global split potential can be written in the form

$$\begin{aligned}k_{\alpha,\beta}(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)&=\ln K_{\alpha,\beta}(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)=\\&=-\sum_{i=1}^{n}\ln c_1f_\alpha(y_i;\theta_{i\alpha})\ln c_2f_\beta(y_i;\theta_{i\beta})\end{aligned} \tag{22}$$

or

$$k_{\alpha,\beta}(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)=-\sum_{i=1}^{n}I_\alpha^f(y_i;\theta_{i\alpha})I_\beta^f(y_i;\theta_{i\beta}) \tag{23}$$

3. Optimization problem and its solution

3.1. General optimization problem

The main principle of M_{split} estimation (Wiśniewski, 2009) assumes that the quantities $\hat{\boldsymbol{\theta}}_\alpha$ and $\hat{\boldsymbol{\theta}}_\beta$ are M_{split} estimates of the parameters $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\beta$, respectively, if only they maximize the global split potential $K_{\alpha,\beta}$ ($\mathbf{y}$; $\boldsymbol{\theta}_\alpha$, $\boldsymbol{\theta}_\beta$). If the potential is referred to the probability families

$$\mathcal{P}_\alpha=\{P_{\boldsymbol{\theta}_\alpha}\succ f_\alpha(\mathbf{y};\boldsymbol{\theta}_\alpha):\boldsymbol{\theta}_\alpha\in\boldsymbol{\Theta}_\alpha\}\quad and\quad \mathcal{P}_\beta=\{P_{\boldsymbol{\theta}_\beta}\succ f_\beta(\mathbf{y};\boldsymbol{\theta}_\beta:\boldsymbol{\theta}_\beta\in\boldsymbol{\Theta}_\beta)\}$$

then the mentioned principle can be formulated as follows

$$\max_{\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta} K_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta) = K_{\alpha,\beta}(\mathbf{y}; \hat{\boldsymbol{\theta}}_\alpha, \hat{\boldsymbol{\theta}}_\beta) \tag{24}$$

Since

$$\max_{\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta} K_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta) = \min_{\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta}[-K_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta)]$$

and

$$\min_{\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta}[-K_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta)] = \min_{\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta}[-k_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta)]$$

then the optimization problem Eq. (24) can be replaced with an equivalent one

$$\min_{\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta}[-k_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta)] = -k_{\alpha,\beta}(\mathbf{y}; \hat{\boldsymbol{\theta}}_\alpha, \hat{\boldsymbol{\theta}}_\beta) \tag{25}$$

Such formulated M_{split} estimation can be regarded as a split ML-estimation. Let the global split potential have more general form

$$k_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta) = -\sum_{i=1}^{n} \rho_\alpha(y_i; \theta_{i\alpha})\rho_\beta(y_i; \theta_{i\beta}) \tag{26}$$

that is not necessarily referred to any probabilistic assumptions. Functions $\rho_\alpha(y_i;\, \theta_{i\alpha})$ and $\rho_\beta(y_i;\, \theta_{i\beta})$ should be at least positive convex and twice differentiable. Introducing the following notations:

$$\varphi(\mathbf{y}; \boldsymbol{\theta}_{\alpha,} \boldsymbol{\theta}_\beta) = -k_{\alpha,\beta}(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta)$$
$$\eta(y_i; \theta_{i\alpha}, \theta_{i\beta}) = \rho_\alpha(y_i; \theta_{i\alpha})\rho_\beta(y_i; \theta_{i\beta})$$

the problem Eq. (25) can be replaced with the following general optimization problem

$$\min_{\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta} \varphi(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta) = \varphi(\mathbf{y}; \hat{\boldsymbol{\theta}}_\alpha, \hat{\boldsymbol{\theta}}_\beta) \tag{27}$$

where

$$\varphi(\mathbf{y}; \boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\beta) = \sum_{i=1}^{n} \eta(y_i; \theta_{i\alpha}, \theta_{i\beta}) = \sum_{i=1}^{n} \rho_\alpha(y_i; \theta_{i\alpha})\rho_\beta(y_i; \theta_{i\beta}) \tag{28}$$

Introducing vectors:

$$\boldsymbol{\rho}_\alpha(\mathbf{y}; \boldsymbol{\theta}_\alpha) = [\rho_\alpha(y_1; \theta_{1\alpha}), \rho(y_2; \theta_{2\alpha}), \cdots, \rho(y_n; \theta_{n\alpha})]^{\mathrm{T}}$$
$$\boldsymbol{\rho}_\beta(\mathbf{y}; \boldsymbol{\theta}_\beta) = [\rho_\beta(y_1; \theta_{1\beta}), \rho(y_2; \theta_{2\beta}), \cdots, \rho(y_n; \theta_{n\beta})]^{\mathrm{T}}$$

the function Eq. (28) is written in the formula

$$\varphi(\mathbf{y};\boldsymbol{\theta}_\alpha,\boldsymbol{\theta}_\beta)=\sum_{i=1}^{n}\rho_\alpha(y_i;\theta_{i\alpha})\rho_\beta(y_i;\theta_{i\beta})=[\boldsymbol{\rho}_\alpha(\mathbf{y};\boldsymbol{\theta}_\alpha)]^{\mathrm{T}}\boldsymbol{\rho}_\beta(\mathbf{y};\boldsymbol{\theta}_\beta) \quad (29)$$

Let us remind that the target function of classic M-estimation is written in the similar form $\varphi(\mathbf{y};\ \boldsymbol{\theta}) = \sum_{i=1}^{n}\rho(y_i;\ \theta_i)$. Thus the estimation based on optimization problem Eq. (27) can be regarded as a split M-estimation (the notation M_{split} is just based on such fact). The function Eq. (29) will be regarded as the general target function of M_{split} estimation.

3.2. Optimization problem and its solution for geodetic observation case

Let

$$\mathbf{v} = \mathbf{y} - \mathbf{AX} \quad (30)$$

be a functional model of an observation vector $\mathbf{y} \in \mathfrak{R}^n$ and let $E(\mathbf{v}) = \mathbf{0}$, $E(\mathbf{y}) = \mathbf{AX}$, ($\mathbf{v}$ – vector of random errors, $\mathbf{A} \in \mathfrak{R}^{n.r}$ – known matrix of coefficients, $\mathbf{X} \in \mathfrak{R}^r$ – vector of unknown parameters). Taking the earlier introduced model $v_i = y_i - \theta_i$ the model Eq. (30) can also be written in the following form

$$v_i = y_i - \theta_i = y_i - \mathbf{a}_{(i,\bullet)}\mathbf{X} \quad (31)$$

where $\mathbf{a}_{(i,\bullet)}$ is i-th row of the matrix $\mathbf{A}$. Let us notice that $\theta_i = \mathbf{a}_{(i,\bullet)}\,\mathbf{X}$ and also $\boldsymbol{\theta} = \mathbf{AX}$. Referring to the earlier presented assumptions, two competitive parameters $\theta_{i\alpha} = \mathbf{a}_{(i,\bullet)}\,\mathbf{X}_\alpha$ and $\theta_{i\beta} = \mathbf{a}_{(i,\bullet)}\,\mathbf{X}_\beta$, therefore, two competitive functional models $v_{i\alpha} = y_i - \mathbf{a}_{(i,\bullet)}\mathbf{X}_\alpha$ and $v_{i\beta} = y_i - \mathbf{a}_{(i,\bullet)}\mathbf{X}_\beta$ respond to each observation y_i. Such natural, as for M_{split} estimation, split of the functional model can be written as follows

$$split(v_i = y_i - \mathbf{a}_{(i,\bullet)}\mathbf{X}) = \begin{cases} v_{i\alpha} = y_i - \mathbf{a}_{(i,\bullet)}\mathbf{X}_\alpha \\ v_{i\beta} = y_i - \mathbf{a}_{(i,\bullet)}\mathbf{X}_\beta \end{cases} \quad (32)$$

or referring to the model $\mathbf{v} = \mathbf{y} - \mathbf{AX}$ (Wiśniewski, 2008)

$$split(\mathbf{v} = \mathbf{y} - \mathbf{AX}) = \begin{cases} \mathbf{v}_\alpha = \mathbf{y} - \mathbf{AX}_\alpha \\ \mathbf{v}_\beta = \mathbf{y} - \mathbf{AX}_\beta \end{cases} \quad (33)$$

The split of the functional model of geodetic observations results in the fact that the optimization problem of M_{split} estimation has two competitive solutions: two estimates of parameters $\hat{\mathbf{X}}_\alpha$ and $\hat{\mathbf{X}}_\beta$ and two competitive residual vectors $\hat{\mathbf{v}}_\alpha$ and $\hat{\mathbf{v}}_\beta$ that respond to the same observation vector $\mathbf{y}$.

Considering the above presented functional model of observations, the general optimization problem Eq. (27) can be replaced with the following one

$$\min_{\mathbf{X}_\alpha, \mathbf{X}_\beta} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = \varphi(\mathbf{y}; \hat{\mathbf{X}}_\alpha, \hat{\mathbf{X}}_\beta) \tag{34}$$

where

$$\begin{aligned} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) &= \sum_{i=1}^{n} \rho_\alpha(y_i; \mathbf{X}_\alpha)\rho_\beta(y_i; \mathbf{X}_\beta) = \\ &= [\rho_\alpha(\mathbf{y}; \mathbf{X}_\alpha)]^{\mathrm{T}} \rho_\beta(\mathbf{y}; \mathbf{X}_\beta) \end{aligned} \tag{35}$$

As it was assumed earlier, the functions $\rho_\alpha(\cdot)$ and $\rho_\beta(\cdot)$ are at least convex and twice differentiable. Thus $\hat{\mathbf{X}}_\alpha$ and $\hat{\mathbf{X}}_\beta$ estimates are a solution of the optimization problem Eq. (34) if only the following gradients of the target function Eq. (35):

$$\begin{aligned} \mathbf{g}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta) &= \frac{\partial}{\partial \mathbf{X}_\alpha} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) \in \Re^r \\ \mathbf{g}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta) &= \frac{\partial}{\partial \mathbf{X}_\beta} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) \in \Re^r \end{aligned}$$

are zero vectors. It means that

$$\mathbf{g}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta)\Big|_{\substack{\mathbf{X}_\alpha = \hat{\mathbf{X}}_\alpha \\ \mathbf{X}_\beta = \hat{\mathbf{X}}_\beta}} = \mathbf{0} \tag{36}$$

$$\mathbf{g}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta)\Big|_{\substack{\mathbf{X}_\alpha = \hat{\mathbf{X}}_\alpha \\ \mathbf{X}_\beta = \hat{\mathbf{X}}_\beta}} = \mathbf{0} \tag{37}$$

at the same time.

By computing the derivative

$$\frac{\partial}{\partial \mathbf{X}_\alpha} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial \mathbf{v}_\alpha}{\partial \mathbf{X}_\alpha} \frac{\partial}{\partial \mathbf{v}_\alpha} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta)$$

one can obtain

$$\frac{\partial}{\partial \mathbf{X}_\alpha} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial \mathbf{v}_\alpha}{\partial \mathbf{X}_\alpha} [\rho_\beta(v_{1\beta}) \frac{\partial \rho_\alpha(v_{1\alpha})}{\partial v_{1\alpha}}, \rho_\beta(v_{2\beta}) \frac{\partial \rho(v_{2\alpha})}{\partial v_{2\alpha}}, \cdots, \rho_\beta(v_{n\beta}) \frac{\partial \rho_\alpha(v_{n\alpha})}{\partial v_{n\alpha}}]^{\mathrm{T}}$$

where (to simplify the notation)

$$\begin{aligned} \boldsymbol{\rho}_\alpha(\mathbf{y}; \mathbf{X}_\alpha) &= [\rho_\alpha(y_1; \mathbf{X}_\alpha), \rho_\alpha(y_2; \mathbf{X}_\alpha), \cdots, \rho(y_n; \mathbf{X}_\alpha)]^{\mathrm{T}} = \\ &= [\rho_\alpha(v_{1\alpha}), \rho(v_{2\alpha}), \cdots, \rho_\alpha(v_{n\alpha})]^{\mathrm{T}} = \boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha) \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\rho}_\beta(\mathbf{y}; \mathbf{X}_\alpha) &= [\rho_\beta(y_1; \mathbf{X}_\beta), \rho_\beta(y_2; \mathbf{X}_\beta), \cdots, \rho(y_n; \mathbf{X}_\beta)]^{\mathrm{T}} = \\ &= [\rho_\beta(v_{1\beta}), \rho(v_{2\beta}), \cdots, \rho_\beta(v_{n\beta})]^{\mathrm{T}} = \boldsymbol{\rho}_\beta(\mathbf{v}_\beta) \end{aligned}$$

By introducing a diagonal matrix

$$\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta)\} = \mathrm{diag}\{\rho_\beta(v_{1\beta}), \rho_\beta(v_{2\beta}), \cdots, \rho_\beta(v_{n\beta})\}$$

furthermore, the terms

$$[\frac{\partial \rho_\alpha(v_{1\alpha})}{\partial v_{1\alpha}}, \frac{\partial \rho_\alpha(v_{2\alpha})}{\partial v_{2\alpha}}, \cdots, \frac{\partial \rho_\alpha(v_{n\alpha})}{\partial v_{n\alpha}}]^{\mathrm{T}} = \frac{\partial \boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)}{\partial \mathbf{v}_\alpha} = \mathbf{g}_{M\alpha}(\mathbf{v}_\alpha)$$

and

$$\frac{\partial \mathbf{v}_\alpha}{\partial \mathbf{X}_\alpha} = \frac{\partial}{\partial \mathbf{X}_\alpha}(\mathbf{y} - \mathbf{A}\mathbf{X}_\alpha) = -\mathbf{A}^{\mathrm{T}}$$

the gradient $\mathbf{g}_\alpha(\hat{\mathbf{X}}_\alpha, \hat{\mathbf{X}}_\beta)$ can be written as

$$\mathbf{g}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial}{\partial \mathbf{X}_\alpha}\varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = -\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta)\}\mathbf{g}_{M\alpha}(\mathbf{v}_\alpha) \tag{38}$$

Analogously, one can write

$$\frac{\partial}{\partial \mathbf{X}_\beta}\varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) =$$

$$= \frac{\partial \mathbf{v}_\beta}{\partial \mathbf{X}_\beta}[\rho_\alpha(v_{1\alpha})\frac{\partial \rho_\beta(v_{1\beta})}{\partial v_{1\beta}}, \rho_\alpha(v_{2\alpha})\frac{\partial \rho_\beta(v_{2\beta})}{\partial v_{2\beta}}, \cdots, \rho_\alpha(v_{n\alpha})\frac{\partial \rho_\beta(v_{n\beta})}{\partial v_{n\beta}}]^{\mathrm{T}} =$$

$$= \frac{\partial \mathbf{v}_\beta}{\partial \mathbf{X}_\beta}\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)\}\frac{\partial \boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)}{\partial \mathbf{v}_\alpha}$$

And since, also in the analogous way,

$$\frac{\partial \mathbf{v}_\beta}{\partial \mathbf{X}_\beta} = \frac{\partial}{\partial \mathbf{X}_\beta}(\mathbf{y} - \mathbf{A}\mathbf{X}_\beta) = -\mathbf{A}^{\mathrm{T}}$$

therefore

$$\mathbf{g}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial}{\partial \mathbf{X}_\beta}\varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = -\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)\}\mathbf{g}_{M\beta}(\mathbf{v}_\beta) \tag{39}$$

where

$$\mathbf{g}_{M\beta}(\mathbf{v}_\beta) = \frac{\partial \boldsymbol{\rho}_\beta(\mathbf{v}_\beta)}{\partial \mathbf{v}_\beta}$$

To solve the optimization problem, i.e. to find such estimates $\hat{\mathbf{X}}_\alpha$ and $\hat{\mathbf{X}}_\beta$ that make

$$\left.\begin{array}{l}\mathbf{g}_\alpha(\hat{\mathbf{X}}_\alpha, \hat{\mathbf{X}}_\beta) = \mathbf{0} \\ \mathbf{g}_\beta(\hat{\mathbf{X}}_\alpha, \hat{\mathbf{X}}_\beta) = \mathbf{0}\end{array}\right\} \quad \Leftrightarrow \quad \min_{\mathbf{X}_\alpha, \mathbf{X}_\beta} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = \varphi(\mathbf{y}; \hat{\mathbf{X}}_\alpha, \hat{\mathbf{X}}_\beta)$$

Newton's method can be applied (e.g. Teunissen, 1990). The iterative process of this method in a case of a particular class of functions $\varphi(\mathbf{y};\ \hat{\mathbf{X}}_\alpha,\ \hat{\mathbf{X}}_\beta)$ was presented in the papers (Wiśniewski, 2008, 2009). The similar form of this process can also be proposed for the general theory of M_{split} estimation that is presented in the present paper. For such purpose, the following Hessians are computed

$$\mathbf{H}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial^2}{\partial \mathbf{X}_\alpha \partial \mathbf{X}_\alpha^{\mathrm{T}}}\varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial}{\partial \mathbf{X}_\alpha^{\mathrm{T}}}\mathbf{g}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta)$$

and

$$\mathbf{H}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial^2}{\partial \mathbf{X}_\beta \partial \mathbf{X}_\beta^{\mathrm{T}}}\varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = \frac{\partial}{\partial \mathbf{X}_\beta^{\mathrm{T}}}\mathbf{g}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta)$$

If

$$\frac{\partial}{\partial \mathbf{X}_\alpha^{\mathrm{T}}}\mathbf{g}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta) = -\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta)\}\frac{\partial \mathbf{g}_{M\alpha}(\mathbf{v}_\alpha)}{\partial \mathbf{v}_\alpha^{\mathrm{T}}}\frac{\partial \mathbf{v}_\alpha}{\partial \mathbf{X}_\alpha^{\mathrm{T}}}$$

and

$$\frac{\partial \mathbf{v}_\alpha}{\partial \mathbf{X}_\alpha^{\mathrm{T}}} = \frac{\partial}{\partial \mathbf{X}_\alpha^{\mathrm{T}}}(\mathbf{y} - \mathbf{A}\mathbf{X}_\alpha) = -\mathbf{A}$$

then Hessian $\mathbf{H}_\alpha(\hat{\mathbf{X}}_\alpha,\ \hat{\mathbf{X}}_\beta)$ can be written as

$$\mathbf{H}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta) = \mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta)\}\frac{\partial \mathbf{g}_{M\alpha}(\mathbf{v}_\alpha)}{\partial \mathbf{v}_\alpha^{\mathrm{T}}}\mathbf{A} \tag{40}$$

The following derivative $\partial \mathbf{g}_{M\alpha}(\mathbf{v}_\alpha)/\partial \mathbf{v}_\alpha^{\mathrm{T}}$ is a diagonal matrix with elements $\partial^2\rho_\alpha(v_{i\alpha})/\partial v_{i\alpha}^2$ ($\partial^2\rho_\alpha(v_{i\alpha})/\partial v_{i\alpha}\partial v_{j\alpha} = 0$ for every $i \neq j$). Let this matrix be denoted as $\mathbf{H}_{M\alpha}(\mathbf{v}_\alpha)$, then the Hessian searched is written in the form

$$\mathbf{H}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta) = \mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta)\}\mathbf{H}_{M\alpha}(\mathbf{v}_\alpha)\mathbf{A} \tag{41}$$

The other Hessian $\mathbf{H}_\beta(\hat{\mathbf{X}}_\alpha,\ \hat{\mathbf{X}}_\beta)$ can be computed in the analogous way. Thus taking

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}_\beta^{\mathrm{T}}}\mathbf{g}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta) &= -\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)\}\frac{\partial \mathbf{g}_{M\beta}(\mathbf{v}_\beta)}{\partial \mathbf{v}_\beta^{\mathrm{T}}}\frac{\partial \mathbf{v}_\beta}{\partial \mathbf{X}_\beta^{\mathrm{T}}} = \\ &= -\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)\}\frac{\partial \mathbf{g}_{M\beta}(\mathbf{v}_\beta)}{\partial \mathbf{v}_\beta^{\mathrm{T}}}\mathbf{A}\end{aligned}$$

and introducing

$$\mathbf{H}_{M\beta}(\mathbf{v}_\beta) = \frac{\partial \mathbf{g}_\beta(\mathbf{v}_\beta)}{\partial \mathbf{v}_\beta^{\mathrm{T}}}$$

one can finally obtain

$$\mathbf{H}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta) = \mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)\}\mathbf{H}_{M\beta}(\mathbf{v}_\beta)\mathbf{A}(42)$$

Let us notify that if the functions $\rho_\alpha(y_i; \mathbf{X}_\alpha) = \rho_\alpha(v_{i\alpha})$ and $\rho_\beta(y_i; \mathbf{X}_\beta) = \rho_\alpha(v_{i\beta})$, $i = 1, ..., n$, are positive and convex then the following multiplications $\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)\}\mathbf{H}_{M\beta}(\mathbf{v}_\beta)$ and $\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta)\}\mathbf{H}_{M\alpha}(\mathbf{v}_\alpha)$ are positive definite matrices. Therefore, Hessians $\mathbf{H}_\alpha(\mathbf{X}_\alpha, \mathbf{X}_\beta)$ and $\mathbf{H}_\beta(\mathbf{X}_\alpha, \mathbf{X}_\beta)$ are also positive definite ones (the sufficient condition is fulfilled).

The iterative process that solves the optimization problem $\min\limits_{\mathbf{X}_\alpha, \mathbf{X}_\beta} \varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = \varphi(\mathbf{y}; \hat{\mathbf{X}}_\alpha, \hat{\mathbf{X}}_\beta)$ by applying Newton's method is two-staged (Wiśniewski, 2008, 2009). When assuming the form of the target function introduced in the present paper, i.e.

$$\varphi(\mathbf{y}; \mathbf{X}_\alpha, \mathbf{X}_\beta) = [\boldsymbol{\rho}_\alpha(\mathbf{y}; \mathbf{X}_\alpha)]^{\mathrm{T}}\boldsymbol{\rho}_\beta(\mathbf{y}; \mathbf{X}_\beta) = [\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha)]^{\mathrm{T}}\boldsymbol{\rho}_\beta(\mathbf{v}_\beta)$$

such process can be written as

$$\left.\begin{aligned} \mathbf{X}_\alpha^j &= \mathbf{X}_\alpha^{j-1} + d\mathbf{X}_\alpha^j \\ \mathbf{X}_\beta^j &= \mathbf{X}_\beta^{j-1} + d\mathbf{X}_\beta^j \end{aligned}\right\}_{j=1, ..., k} \tag{43}$$

where

$$\left.\begin{aligned} d\mathbf{X}_\alpha^j &= -\{\mathbf{H}_\alpha(\mathbf{X}_\alpha^{j-1}, \mathbf{X}_\beta^{j-1})\}^{-1}\mathbf{g}_\alpha(\mathbf{X}_\alpha^{j-1}, \mathbf{X}_\beta^{j-1}) = \\ &= \{\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta^{j-1})\}\mathbf{H}_{M\alpha}(\mathbf{v}_\alpha^{j-1})\mathbf{A}\}^{-1}\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\beta(\mathbf{v}_\beta^{j-1})\}\mathbf{g}_{M\alpha}(\mathbf{v}_\alpha^{j-1}) \\ d\mathbf{X}_\beta^j &= -\{\mathbf{H}_\beta(\mathbf{X}_\alpha^j, \mathbf{X}_\beta^{j-1})\}^{-1}\mathbf{g}_\beta(\mathbf{X}_\alpha^j, \mathbf{X}_\beta^{j-1}) = \\ &= \{\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha^j)\}\mathbf{H}_{M\beta}(\mathbf{v}_\beta^{j-1})\mathbf{A}\}^{-1}\mathbf{A}^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{\rho}_\alpha(\mathbf{v}_\alpha^j)\}\mathbf{g}_{M\beta}(\mathbf{v}_\beta^{j-1}) \end{aligned}\right\}_{j=1, ..., k} \tag{44}$$

The iterative process ends when $\hat{\mathbf{X}}_\alpha = \mathbf{X}_\alpha^j = \mathbf{X}_\alpha^{j-1}$ and $\hat{\mathbf{X}}_\beta = \mathbf{X}_\beta^j = \mathbf{X}_\beta^{j-1}$. The estimates $\hat{\mathbf{v}}_\alpha = \mathbf{y} - \mathbf{A}\hat{\mathbf{X}}_\alpha$ and $\hat{\mathbf{v}}_\beta = \mathbf{y} - \mathbf{A}\hat{\mathbf{X}}_\beta$ that are the competitive vectors of residuals assigned to the same observation vector $\mathbf{y}$ can be computed on the base of the M_{split} estimators $\hat{\mathbf{X}}_\alpha$ and $\hat{\mathbf{X}}_\beta$, respectively.

4. Conclusions

The theory presented in the present paper supplements the theory of M_{split} estimation. It is especially because the assumptions concerning local and global split potentials are extended. It is possible by applying the f-information introduced here. The assumption extension is possible since the f-information is based on probability density functions.

M_{split} estimates are such quantities that maximize the global split potential. Therefore, two competitive probability density functions concerning the same observation vector should be assumed to formulate and to solve the optimization problem

Eq. (24) (or equivalently Eq. (25)). Some statistical methods can be applied to choose such suitable competitive functions (e.g. Romanowski, 1979; Wiśniewski, 1985, 1987, 1996). It is also possible to base such choice on some general arbitrary and theoretical assumptions. For example, Wiśniewski (2009) assumed that $\mathcal{P}_\alpha$ and $\mathcal{P}_\beta$ are the families of normal probability distributions (the same assumption will be also applied in the second part of the present paper).

The general form of the M_{split} target function Eq. (29) can be proposed when f-information is replaced with at least convex and twice differentiable functions. Such functions can be chosen arbitrarily thus one can suppose that the class of M_{split} estimator is broad and contains estimates of different properties.

The optimization problem of M_{split} estimation (27) is a general one. It can be changed to the Eq. (34) when it is applied to geodetic problems. Then two competitive M_{split} estimates $\hat{\mathbf{X}}_\alpha$ and $\hat{\mathbf{X}}_\alpha$ are its solutions. This two estimates refer to the same observation vector $\mathbf{y}$ (if only the original functional model $\mathbf{v} = \mathbf{y} - \mathbf{AX}$ is split into the two new ones $\mathbf{v}_\alpha = \mathbf{y} - \mathbf{AX}_\alpha$ and $\mathbf{v}_\beta = \mathbf{y} - \mathbf{AX}_\beta$). For that reason, also two competitive residual vectors $\hat{\mathbf{v}}_\alpha$ and $\hat{\mathbf{v}}_\beta$ refer to the same observation vector.

Functions $\rho_\alpha(y_i; \mathbf{X}_\alpha)$ and $\rho_\beta(y_i; \mathbf{X}_\beta)$ proposed in the present paper have such general properties that the optimization problem can be solved applying the Newton method. The iterative process is described and the necessary forms for gradients and Hessians are derived (see, Eqs. (37)-(41)). The procedure proposed here is two-staged and the second stage is party based on the results of the first one.

The present part of the paper is a theoretical one and can be regarded as a generalization of the earlier published theory of M_{split} estimation (Wiśniewski, 2009). The next part, which is also strictly referred to the above mentioned paper, presents and develops theory of a squared M_{split} estimation, i.e. such M_{split} estimation where functions $\rho_\alpha(\cdot)$ and $\rho_\beta(\cdot)$ are squared ones. This kind of estimation seems to be very useful from practical point of view (as for the present stage of M_{split} estimation development). Thus the next part presents some numerical examples that illustrate the main properties of this kind of M_{split} estimation. Those examples together with the earlier published ones (Wiśniewski, 2008, 2009) point at the future application of the newly elaborated method of estimation.

Acknowledgments

The research is supported by the Polish Ministry of Science and Higher Education under the grant "New methods of estimation of parameters in a functional model of geodetic observations (M_{split} estimation)", No. N N526 255 734.

References

Elderton W.P., (1953): *Frequency curves and correlation*, Cambridge University Press.

Götzelmann M., Keller W., Reubelt T., (2006): *Gross error compensation for gravity field analysis based on kinematic orbit data*, Journal of Geodesy, Vol. 80, pp. 184-198.

Grodecki J., (1999): *Generalized maximum-likelihood estimation of variance components with inverted gamma prior*, Journal of Geodesy, Vol. 73, pp. 367-374.

Gui Q., Zhang J., (1998): *Robust biased estimation and its applications in geodetic adjustment*, Journal of Geodesy, Vol. 72, pp. 430-435.

Hampel F.R., (1974): *The influence curve and its role in robust estimation*, Journal of the American Statistical Association, Vol. 69, pp. 383-397.

Hampel F.R, Ronchetti E.M., Rousseeuw P.J., Stahel W.A., (1986): *Robust statistics. The approach based on influence functions*, John Wiley & Sons, New York.

Huang Y., Mertikas S.P., (1995): *On the design of robust regression estimators*, Manuscripta Geodaetica, Vol. 20, pp. 145-160.

Huber P.J., (1964): *Robust estimation of location parameter*, Annals of Mathematical Statistics, Vol. 43 (4), pp. 1041-1067.

Huber P.J., (1981): *Robust statistics*, John Wiley & Sons, New York.

Jones G.A., Jones J.M., (2000): *Information and coding theory*, Springer (Springer Undergraduate Mathematics Series).

Kadaj R., (1984): *Die Methode der besten Alternative: ein Ausgleichungsprinzip für Beobachtungssysteme*, Zeitschrift für Vermessungswesen, Vol. 109, No 6, pp. 301-308.

Kadaj R., (1988): *Eine verallgemeinerte Klasse von Schatzverfahren mit praktischen Anwendungen*, Zeitschrift für Vermessungswesen, Vol. 113, No 4, pp. 157-166.

Koch K.R., (1986:) *Maximum Likelihood estimate of variance components*, Bulletin Géodésique, Vol. 60, pp. 329-338.

Koch K.R., (1996): *Robuste Parameterschätzung*, Allgemeine Vermessungs Nachrichten, Vol. 103, No 11, pp. 1-18.

Koch K.R., Yang Y., (1998): *Robust Kalman filter for rank deficient observation models*, Journal of Geodesy, Vol. 72, pp. 436-441.

Krarup T., Kubik K., (1983): *The Danish Method; experience and philosophy*, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Reihe A, Heft Nr 7, München, pp. 131-134.

Prószyński W., (1997): *Measuring the robust potential of the least-squares estimation: geodetic illustration*, Journal of Geodesy, Vol. 71, pp. 652-659.

Prószyński W., (2000): *On outlier-hiding effects in specific Gauss-Markov models: geodetic examples*, Journal of Geodesy, Vol. 74, pp. 581-589.

Rousseeuw P.J., (1984): *Least median of squares regression*, Journal of the American Statistical Association, Vol. 79, pp. 871-880.

Romanowski M., (1979): *Random errors in observations and the influence of modulation on their distribution*, Verlag Konrad Wittwer, Stuttgart.

Serfling R.J., (1980): *Approximation theorems of mathematical statistics*, John Wiley & Sons (Polish Edition 1991, PWN, Warszawa).

Teunissen P.J.G., (1990): *Nonlinear least squares*, Manuscripta Geodaetica, Vol. 15, pp. 137-150.

Wiśniewski Z., (1985): *The effect of the asymmetry of geodetic observation error distribution on the results of adjustment by the least squares method*, Geodesy and Cartography, Vol. 34, No 1, pp. 11-21.

Wiśniewski Z., (1987): *Method of geodetic network adjustment in extended to probabilistic measurement error properties*, Scientific Bulletin of the Staszic Academy of Mining and Metallurgy, Geodesy, Vol. 95, pp. 73-88.

Wiśniewski Z., (1993): *Robustness properties of the RP method*, Geodesy and Cartography, t. XLII (2), pp. 135-151.

Wiśniewski Z., (1996): *Estimation of third and fourth order central moments of measurement errors from sums of powers of least squares adjustment residuals*, Journal of Geodesy, Vol. 70, pp. 256-262.

Wiśniewski Z. (1999): *Concept of robust estimation of variance coefficient (VR-estimation)*, Bollettino di Geodesia e Scienze Affini, Vol. 58, No 3, pp. 291-310.

Wiśniewski Z., (2008): *Split estimation of parameters in functional geodetic models*, Technical Sciences, Vol. 11, pp. 202-212.
Wiśniewski Z., (2009): *Estimation of parameters in a split functional model of geodetic observations (M_{split} estimation)*, Journal of Geodesy, Vol. 83, pp. 105-120.
Xu P., (1989): *On robust estimation with correlated observations*, Bulletin Géodésique, Vol. 63, pp. 237-252.
Xu P., (2005): *Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness*, Journal of Geodesy, Vol. 79, pp.146-159.
Yang Y., (1991): *Robust Bayesian estimation*, Bulletin Géodésique, Vol. 65, pp. 145-150.
Yang Y., (1992): *Robustifying collocation*, Manuscripta Geodaetica, Vol. 17, pp. 21-28.
Yang Y., (1994): *Robust estimation for dependent observations*, Manuscripta Geodaetica, Vol. 19, pp. 10-17.
Yang Y., (1997): *Estimators of covariance matrix at robust estimation based on influence functions*, Zeitschrift für Vermessungswesen, Vol. 122, No 4, pp. 166-174.
Yang Y., (1999): *Robust estimation of geodetic datum transformation*, Journal of Geodesy, Vol. 73, pp. 268-274.
Yang Y., He H., Xu G., (2001): *Adaptively robust filtering for kinematic geodetic positioning*, Journal of Geodesy, Vol. 75, pp. 106-109.
Yang Y., Song L., Xu T., (2002): *Robust estimation for correlated observations based on bifactor equivalent weights*, Journal of Geodesy, Vol. 76, pp. 353-358.
Yu Z.C., (1996): *A universal formula of maximum likelihood estimation of variance-covariance components*, Journal of Geodesy, Vol. 70, pp. 233-240.
Zhong D., (1997): *Robust estimation and optimal selection of polynomial parameters for the interpolation of GPS geoid heights*, Journal of Geodesy, Vol. 71, pp. 552-561.

M_{split} estymacja. Część I. Podstawy teoretyczne

Zbigniew Wiśniewski

Instytut Geodezji
Uniwersytet Warmińsko-Mazurski
ul. Oczapowskiego 1, 10-957 Olsztyn
e-mail: zbyszek@uwm.edu.pl

Streszczenie

Niniejsza praca składa się z dwóch części. W pierwszej z nich, w nawiązaniu do wcześniejszej pracy autora (Wiśniewski, 2009) przedstawiono teoretyczne podstawy M_{split} estymacji. W stosunku do cytowanej pracy, tutaj bardziej szczegółowo omówiono założenia o charakterze probabilistycznym. Wprowadzono także pojęcie f-informacji co pozwoliło na zaproponowanie bardziej ogólnej formy potencjału rozszczepienia. Podstawową treścią tej części pracy jest uogólnienie funkcji celu M_{split} estymacji. Dla tej funkcji oraz w odniesieniu do modelu obserwacji geodezyjnych, ustalono problem optymalizacyjny oraz przedstawiono sposób jego rozwiązania.

W drugiej części pracy, także w nawiązaniu do cytowanej pracy autora, przedstawiono pewien szczególny przypadek M_{split} estymacji nazwany kwadratową M_{split} estymacją. Rozwinięto teorię tej wersji M_{split} estymacji oraz przedstawiono kilka przykładów numerycznych wskazujących na jej podstawowe własności oraz możliwe obszary zastosowania.